

# Duality and Nonlinear Graph Laplacians

Eric J. Friedman  
International Computer Science Institute  
ejf@icsi.berkeley.edu

Adam S. Landsberg  
W. M. Keck Science Department of Claremont  
McKenna, Pitzer, and Scripps Colleges  
ALandsberg@kecksci.claremont.edu

## ABSTRACT

We present an iterative algorithm for solving a class of nonlinear Laplacian system of equations in  $\tilde{O}(k^2 m \log(kn/\epsilon))$  iterations, where  $k$  is a measure of nonlinearity,  $n$  is the number of variables,  $m$  is the number of nonzero entries in the graph Laplacian  $L$ ,  $\epsilon$  is the solution accuracy and  $\tilde{O}()$  neglects (non-leading) logarithmic terms. This algorithm is a natural nonlinear extension of the one by of Kelner et. al., which solves a linear Laplacian system of equations in nearly linear time. Unlike the linear case, in the nonlinear case each iteration takes  $\tilde{O}(n)$  time so the total running time is  $\tilde{O}(k^2 mn \log(kn/\epsilon))$ . For sparse graphs where  $m = O(n)$  and fixed  $k$  this nonlinear algorithm is  $\tilde{O}(n^2 \log(n/\epsilon))$  which is slightly faster than standard methods for solving linear equations, which require approximately  $O(n^{2.38})$  time. Our analysis relies on the construction of a nonlinear “energy function” and a nonlinear extension of the duality analysis of Kelner et. al to the nonlinear case without any explicit references to spectral analysis or electrical flows. These new insights and results provide tools for more general extensions to spectral theory and nonlinear applications.

## General Terms

Graph Laplacian, Spectral analysis, Nonlinear equations

## 1. INTRODUCTION

The goal of solving linear systems of equations in nearly linear time, spearheaded by Vaidya [10], has propelled a revolution in spectral graph theory and several related fields. In addition to many important results about graph sparsification (see [6, 11] for overviews and references) this theory has led to two main types of methods of solving sets of symmetrical diagonally dominant linear equations in nearly linear time – the first using graph sparsification (and ultra sparsifiers) (e.g., [8]) combined with classical iterative methods, and the second which relies on a underlying low stretch spanning tree to provide an algorithm that relies on cycle updates for “electrical flows” in the underlying graph

[3]. In addition to having important applications in matrix computations, these results have led to a large number theoretical results, reducing the complexity of several important graphical algorithms.

The key insights underlying these approaches are the reduction of a set of symmetrical diagonally dominant linear equations to a system of equations  $Lx = b$  arising from a graph Laplacian,  $L$ . Previous analyses (e.g., [8]) rely on the linearity of the underlying system of equations and the use of a carefully chosen spanning tree on the underlying graph. Thus, one might expect that linearity is required for the analyses; however, as we will show, linearity is not necessary and there appears to be a more general structure underlying these approaches which allows a certain degree of nonlinearity. More precisely, we show that one can generalize the graph Laplacian in a nonlinear manner and apply a nonlinear extension of the Algorithm in Kelner et al. [3] to solve them in  $\tilde{O}(k^2 m \log(kn/\epsilon))$  iterations and  $\tilde{O}(k^2 mn \log(kn/\epsilon))$  time. Our analysis relies on the construction of an “nonlinear energy function” and extension of their duality analysis to nonlinear Laplacian systems of equations. While providing new insights into the linear problem, this analysis shows that one can generalize graph Laplacians without losing much of the underlying structure and suggests further nonlinear generalizations.

## 2. RELATED WORK

The idea of using a spanning tree to precondition a set of linear equations comes from an unpublished presentation by Vaidya [10]. That paper led to a large body of work over many years that led to Spielman and Teng’s seminal paper [8] which showed the theoretical existence of nearly linear time algorithms for systems of equations, using very sophisticated machinery [4, 5, 7, 9].

Our algorithm is based on Kelner et. al’s algorithm [3], which dramatically reduces the amount of machinery required. Its origins are clearly implicitly based on the more sophisticated machinery in the earlier papers, such as low-stretch spanning trees (e.g., [1]) and ultra-sparsifiers (e.g., [8]), but does not require them directly.

## 3. GENERALIZED MODEL AND RELAXATION

We begin with an undirected weighted graph  $G = (V, E, w)$  with node set  $V$ , edges  $E$  and weights  $w_{ij} > 0$  where  $w_{ij} = w_{ji}$ . The graph Laplacian of  $G$  is given by  $L$  where  $L_{ij} =$

$-w_{ij}$  for  $i \neq j$  and  $L_{ii} = -\sum_{j \neq i} w_{ij}$ . Given a vector  $b \in \mathbb{R}^n$  a Laplacian system of equations is simply  $Lx = b$ . Such a system always has a solution if  $\sum_i b_i = 0$ , which we denote  $x^*$  and is unique up to an additive constant. Throughout this paper whenever we talk about solutions, we will ignore the multiplicity of solutions due to the nullspace of  $L$ . Note that for notational clarity we will always assume that  $G$  is connected. In the case that  $G$  is not connected one can simply solve on each connected component separately.

Let  $N(i)$  be the neighbors of node  $i$  in the graph  $G$ . It is well known that one can express the action of the Laplacian as

$$(Lx)_i = \sum_{j \in N(i)} w_{ij}(x_i - x_j),$$

which only depends on  $x_i - x_j$  for  $(i, j) \in E$ . In the following we will consider a generalization of this to non-linear graph Laplacians where

$$(Lx)_i = \sum_{j \in N(i)} w_{ij} h_{ij}(x_i - x_j) \quad (*)$$

where the nonlinear function  $h_{ij}(\cdot)$  is anti-symmetric, satisfies  $h_{ij}(-v) = -h_{ij}(v)$ , is continuous, the derivative  $h'_{ij}(v)$  is left-continuous,  $vh'_{ij}(v)$  is strictly increasing and  $1/k \leq h'_{ij}(v) < k$  for a chosen  $k \geq 1$ .

Note that the running time of our algorithm will depend on  $k$  and the computational complexity of  $h_{ij}(\cdot)$ . In order to simplify the analysis with respect to  $h_{ij}(\cdot)$ , we assume that we have constant time oracles for computing  $h_{ij}(\cdot)$ ,  $h_{ij}^{-1}(\cdot)$ , and  $h'_{ij}(\cdot)$ . The more natural version where these oracles require  $O(\log(1/\epsilon))$  computations, would only increase our running times by  $O(\log(1/\epsilon))$  factors, where  $\epsilon$  is the accuracy of the solution.

A simple example of this model arises when  $h'_{ij}(v) = 1/k$  for  $|v| \leq 1$  and  $h'_{ij}(v) = 1$  otherwise. This example arises when the cost is small for small flows, in contrast to the linear case where the “cost” of a flow between edges is linear in the flow. More generally, we can choose  $h'_{ij}(v)$  to be constant except at a discrete set of points where it increases, corresponding to a piecewise linear  $h_{ij}(v)$  with discrete increases in slope, bounded by slopes  $1/k$  and  $k$ . Surprisingly, one can also use an  $h_{ij}(v)$  that is not convex in the positive quadrant. For example the function given by  $h_{ij}(v) = v + \arctan(v)$  satisfies the requirements and is not convex in the positive quadrant, leading to “economies of scale” where the incremental cost can decrease as a flow increases.

Following Kelner et. al [3] we can introduce new variables,  $g_{ij}$ , which for generalized Laplacians allow us to rewrite  $Lx = b$  as:

$$\forall i \in V : \sum_{j \in N(i)} w_{ij} h_{ij}(g_{ij}) = b_i \quad (BC_i)$$

$$\forall (i, j) \in E : g_{ij} = x_i - x_j \quad (PC_i)$$

which decomposes the problem in a natural way into b-constraints  $(BC_i)$  and p-constraints  $(PC_i)$ , where  $x$  is a ‘potential function’ for  $g$ . We say that  $g$  is b-feasible if  $g$  satisfies the b-constraints and p-feasible for p-constraints.

Formally, we will only have a single variable for each edge, written  $g_e$  for  $e \in E$  but for notational convenience we will write  $g_{ij} = -g_{ji}$  where one of these two variables as canonical and the other is simply a notational convenience. For example if  $\sum_{(i,j) \in E} g_{ij}$  corresponds to a sum where each edge appears once, while in  $\sum_{i \in V} \sum_{j \in N(i)} g_{ij}$  each  $g_e$  appears twice.

Note that our variables differ from those in Kelner et. al [3] and most of the literature in the linear case where in the electrical model our  $g_{ij}$  variables correspond to the voltage drop between nodes  $i$  and  $j$  while their  $h_{ij}$  corresponds to the current flow between those nodes. Algebraically,  $g_{ij} = h_{ij}/w_{ij}$ .

## 4. CYCLE UPDATES

A key idea in the algorithm by Kelner et al [3] is to update the  $g_{ij}$ ’s using cycles to maintain b-feasibility. Let  $C$  be an oriented set of edges in  $G$  that forms a cycle. Let  $\alpha(C, t)$  be defined such that for all  $(i, j) \in C$ ,

$$w_{ij} h_{ij}(g_{ij} + \alpha_{ij}(C, t)) = w_{ij} h_{ij}(g_{ij}) + t,$$

and  $\alpha_{ij}(C, t) = 0$  for  $(i, j) \notin C$  and note that  $\alpha(C, 0) = 0$  by construction. It can be easily checked that this procedure maintains b-feasibility for any value of  $t$ , since the increase from an incoming edge of the cycle through any node is canceled by the decrease in the sum from the outgoing edge that follows it. When  $h_{ij}(\cdot)$  is the identity,  $\alpha_{ij}(C, t) = t/w_{ij}$  as in the linear analysis.

The key idea in the linear analysis is to iteratively update cycles by updating cycles using  $\alpha(C, t)$  in order to approach p-feasibility while always maintaining b-feasibility. The analysis [3] relies on duality related to “electrical systems” with an energy function. In the nonlinear case we can also define an analogous energy function

$$\Phi(g) = \sum_{ij} \phi_{ij}(g_{ij})$$

where  $g$  is the vector of  $g_{ij}$ ’s and

$$\phi_{ij}(g_{ij}) = w_{ij} \int_0^{g_{ij}} s h'_{ij}(s) ds.$$

In the linear case where  $h_{ij}(v)$  is the identity, this reduces to the standard energy function of [3], after a change of variables.

The key insight in our analysis is that for any cycle  $C$  the “cycle adjustment”,  $\alpha(C, t)$  that minimizes  $\Phi(g)$  satisfies p-feasibility around the cycle.

LEMMA 1. *If for some cycle  $C$*

$$t^* = \operatorname{argmin}_t \Phi(g + \alpha(C, t))$$

*then*

$$\sum_{(i,j) \in C} (g_{ij} + \alpha_{ij}(C, t^*)) = 0.$$

PROOF. To see this note that  $\Phi(g)$  is convex by the defi-

dition of the  $h$ 's and that the derivative is

$$\begin{aligned}\frac{d\Phi(g + \alpha(C, t))}{dt} &= \sum_{(i,j) \in C} \frac{d\phi_{ij}(g_{ij} + \alpha_{ij}(C, t))}{dt} \\ &= \sum_{(i,j) \in C} \phi'_{ij}(g_{ij} + \alpha_{ij}(C, t)) \frac{d\alpha_{ij}(C, t)}{dt}.\end{aligned}$$

Inserting the definition of  $\phi_{ij}$  into the right hand side of the equation yields

$$= \sum_{(i,j) \in C} w_{ij}(g_{ij} + \alpha_{ij}(C, t)) h'_{ij}(g_{ij} + \alpha_{ij}(C, t)) \frac{d\alpha_{ij}(C, t)}{dt}.$$

Implicitly differentiating the defining equation for  $\alpha(C, t)$  with respect to  $t$  gives

$$w_{ij} h'_{ij}(g_{ij} + \alpha_{ij}(C, t)) \frac{d\alpha_{ij}(C, t)}{dt} = 1, \quad (**)$$

which combined with the previous equation and the definition of  $t^*$  gives the desired result after setting the derivative to 0.  $\square$

This lemma implies our first theorem:

**THEOREM 1.** *If  $g$  is a  $b$ -feasible flow that minimizes  $\Phi(g)$  over all "cycle updates" in  $G$ , then  $g$  is also  $p$ -feasible.*

**PROOF.** Clearly if there are no cycle updates that decrease the energy then by the preceding Lemma the flows around those cycles must sum to 0.  $\square$

We next derive a quantitative measure of the gain in energy from a cycle update. This will depend on both the initial flow around the cycle

$$G_C = \sum_{(i,j) \in C} g_{ij}$$

and the harmonic sum of the  $w$ 's around the cycle

$$W_C = \left( \sum_{(i,j) \in C} 1/w_{ij} \right)^{-1}.$$

**LEMMA 2.** *For a cycle  $C$  the decrease in energy from a cycle update satisfies:*

$$\Phi(g) - \Phi(g + \alpha(C, t^*)) \geq W_C G_C^2 / (2k).$$

**PROOF.** We assume that  $G_C \leq 0$ . (If not then simply traverse the cycle in the opposite direction.) As seen in the proof in Lemma 1

$$\frac{d\Phi(g + \alpha(C, t))}{dt} = \sum_{(i,j) \in C} (g_{ij} + \alpha_{ij}(C, t)).$$

Integrating from  $t = t^*$  to  $t = 0$  yields

$$\Phi(g) - \Phi(g + \alpha(C, t^*)) = \int_{t^*}^0 (G_C + \sum_{(i,j) \in C} \alpha_{ij}(C, t)) dt \quad (**).$$

Note that the integrand is 0 at  $t = t^*$ ,  $G_C$  at  $t = 0$  and is strictly decreasing over the range of integration. Using (\*\*) we see that

$$\frac{d\alpha_{ij}(C, t)}{dt} = (w_{ij} h'_{ij}(g_{ij} + \alpha_{ij}(C, t)))^{-1} \leq k/w_{ij}$$

Summing over  $C$  and integrating over  $t$  yields

$$\sum_{(i,j) \in C} \alpha_{ij}(C, t) \leq kt/W_C$$

Since  $G_C + \sum_{(i,j) \in C} \alpha_{ij}(C, t^*) = 0$  this implies that  $t^* \geq -W_C G_C / k$ . Combining this with (\*\*) yields

$$\Phi(g) - \Phi(g + \alpha(C, t^*)) \geq \int_{-W_C G_C / k}^0 (G_C + kt/W_C) dt$$

and integrating completes the proof.  $\square$

To compute the cycle update we can use Newton's method on the derivative  $\Phi(g + \alpha(C, t))$ , written in the form.

$$G_C + \sum_{(i,j) \in C} \alpha_{ij}(C, t).$$

Note that the sum is bounded by  $(t/W_C)/k$  and  $(t/W_C)k$  so we can use a binary search on the interval  $[0, (t/W_C)k]$ . Since the function is convex in  $t$ , for any  $t \in [t^*/4, 3t^*/4]$  the value of the function will be reduced by at least a factor of 2. This will require  $O(\log(k^2))$  iterations so the total time required is  $O(n \log(k))$  since each evaluation of the sum requires  $O(n)$  time.

Thus we have proven an approximate version of the previous lemma.

**LEMMA 3.** *One can compute an approximate cycle update in  $O(n \log(k))$  such that for any cycle  $C$  the decrease in energy satisfies:*

$$\Phi(g) - \Phi(g + \alpha(C, t^*)) \geq W_C G_C^2 / (4k).$$

## 5. SPANNING TREES

Given a spanning tree,  $T \subseteq E$  one can analyze a simple set of cycles that span the cycle space of all cycles [2]. Given an edge  $(i, j) \in E \setminus T$  define the tree path  $P_{(i,j)}$  to be the minimal oriented set of edges from node  $j$  to node  $i$  and let the tree cycle to be  $C_{(i,j)} = P_{(i,j)} \cup (i, j)$ . Thus, instead of minimizing  $\Phi(g)$  over all cycles one only needs to minimize it over each of the tree cycles.

As in [3], we will choose edges  $e \in E \setminus T$  at random with probability  $p_e$  and then minimize  $\Phi(g)$  over the tree cycle  $C_e$ . The main difference here is that in the linear case one can minimize the energy over a cycle analytically while in the nonlinear case one needs to rely on an iterative method, which we described earlier.

For later use, we define the tree-values of  $x$ , denoted  $\hat{x}(T, g)$  to be those which satisfy  $g_{ij} = x_i - x_j$  on the tree. These are easily computed, up to an additive constant, using a breadth first search from any leaf of the spanning tree of  $G$ .

## 6. LAGRANGIAN DUALITY

In order to analyze the progress made by each iteration of our algorithm we need an estimate of the distance to optimality. To do this we compute the Lagrangian dual for our minimization problem.

First recall that our primal problem is to compute

$$\min_g \Phi(g) \quad \text{s.t.} \quad \sum_{j \in N(i)} w_{ij} h_{ij}(g_{ij}) = b_i, \quad \forall i \in V$$

The dual function is given by

$$\Theta(x) = \min_g \left\{ \Phi(g) - \sum_{i \in V} x_i \left[ \sum_{j \in N(i)} w_{ij} h_{ij}(g_{ij}) - b_i \right] \right\}$$

where  $x$  are the dual variables, and as we will see correspond to the  $x$  in our original problem. (As a reminder, we only consider one of  $g_{ij}, g_{ji}$  as a variable, say  $g_{ij}$  and implicitly replace all occurrences of  $g_{ji}$  with  $-g_{ij}$ .) The first order condition for this minimization when differentiating by  $g_{ij}$  is:

$$0 = \phi'_{ij}(g_{ij}) - (x_i - x_j) w_{ij} h'_{ij}(g_{ij})$$

Using the definition of  $\phi_{ij}$  yields:

$$0 = w_{ij} h'_{ij}(g_{ij}) g_{ij} - (x_i - x_j) w_{ij} h'_{ij}(g_{ij})$$

which reduces to  $g_{ij} = x_i - x_j$ . Plugging this into  $\Theta(x)$  yields

$$\Theta(x) = \sum_{(i,j) \in E} \phi_{ij}(x_i - x_j) - \sum_{i \in V} x_i \left[ \sum_{j \in N(i)} w_{ij} h_{ij}(x_i - x_j) - b_i \right]$$

To simplify this we define  $\tilde{g}_{ij} = x_i - x_j$  and use the b-feasibility of  $g_{ij}$  to write this as

$$\Theta(x) = \sum_{(i,j) \in E} \phi_{ij}(\tilde{g}_{ij}) - \sum_{(i,j) \in E} w_{ij} \tilde{g}_{ij} (h_{ij}(\tilde{g}_{ij}) - h_{ij}(g_{ij})).$$

Using this we can approximate the distance to optimality since  $\Phi(g) \geq \Theta(x)$  for any b-feasible  $g$  and  $\Phi(g^*) = \Theta(x^*)$  where  $g^*, x^*$  are the optimizers of their respective problems. Thus, following Kelner et. al if we define

$$GAP(g) = \Phi(g) - \Phi(g^*)$$

then for any b-feasible  $g$  we have

$$GAP(g) \leq \Phi(g) - \Theta(x)$$

We can use the tree values for  $x$  to provide this estimate yielding

$$GAP(g) \leq TGAP = \sum_{(i,j) \in E \setminus T} TGAP_{ij}(g)$$

where we define TGAP by

$$TGAP_{ij}(g) = \phi_{ij}(g_{ij}) - \phi_{ij}(\hat{x}_i - \hat{x}_j) + w_{ij}(\hat{x}_i - \hat{x}_j)(h_{ij}(\hat{x}_i - \hat{x}_j) - h_{ij}(g_{ij}))$$

where the sums in the gap formula need only be over non-tree edges since  $x_i - x_j = g_{ij}$  for all  $(i, j) \in T$ .

LEMMA 4. *Given a b-feasible  $g$ ,*

$$GAP(g) \leq \sum_{(i,j) \in E \setminus T} w_{ij} G_{C_{ij}}^2 k/2$$

where  $C_{ij}$  is the tree cycle passing through edge  $(i, j)$ .

PROOF. Since  $GAP \leq TGAP$  we can focus on TGAP,

$$\frac{\partial TGAP_{ij}(g)}{\partial g_{ij}} = \phi'_{ij}(g_{ij}) - w_{ij}(\hat{x}_i - \hat{x}_j) h'_{ij}(g_{ij})$$

substituting for  $\phi'_{ij}$  gives

$$\frac{\partial TGAP_{ij}(g)}{\partial g_{ij}} = w_{ij}(g_{ij} - (\hat{x}_i - \hat{x}_j)) h'_{ij}(g_{ij}).$$

Integrating yields

$$TGAP_{ij}(g) \leq \int_{(\hat{x}_i - \hat{x}_j)}^{g_{ij}} w_{ij}(s_{ij} - (\hat{x}_i - \hat{x}_j)) k \, ds_{ij}$$

which after noting that  $G_{C_{ij}} = (g_{ij} - (\hat{x}_i - \hat{x}_j))$  yields

$$TGAP_{ij}(g) \leq w_{ij} G_{C_{ij}}^2 k$$

which implies the result after summing.  $\square$

## 7. TREE CONDITION NUMBERS

In order to guarantee convergence in the linear case, Kelner et. al. [3] require that the underlying spanning tree have nice properties. As we will show the same condition suffices for the nonlinear case as well.

Define the tree condition number of spanning tree  $T$  to be

$$\tau(T) = \sum_{(i,j) \in E \setminus T} w_{ij} / W_{C_{ij}}$$

where  $C_{ij}$  is the tree cycle over edge  $(i, j)$ . Now, as discussed in Kelner et. al, the tree condition number is closely related to the stretch of a tree, a well studied topic. The stretch of a tree  $T$  is defined to be  $st(T) = \sum_{(i,j) \in E} st(i, j)$  where

$$st(i, j) = w_{ij} \sum_{(r,s) \in P(i,j)} 1/w_{rs}$$

and it is easy to show that  $\tau(T) = st(T) + m - 2n + 2$ .

Thus, we can use the following result due to Abraham et. al [1] to construct a tree with a good condition number

THEOREM 2 (ABRAHAM ET. AL [1]). *In  $O(m \log n \log \log n)$  time we can compute a spanning tree  $T$  with total stretch  $st(T) = O(m \log n \log \log n)$ .*

## 8. LINEARIZATION

For static analyses we can linearize the Laplacian at some  $\hat{g}$  by considering the adjusted weights,  $\hat{w}_{ij}^{\hat{g}} = w_{ij} h_{ij}(\hat{g}_{ij}) / \hat{g}_{ij}$ , creating a new Laplacian  $\hat{L}^{\hat{g}}$ . Then we can write

$$(\hat{L}^{\hat{g}} x)_i = \sum_{j \in N(i)} \hat{w}_{ij}^{\hat{g}} (x_i - x_j).$$

Next we define the linearized energy to be the energy of the linearized Laplacian:

$$\hat{\xi}^{\hat{g}}(g) = \sum_{(i,j) \in E} \hat{w}_{ij}^{\hat{g}} g_{ij}^2 / 2$$

and it is easy to see that  $\hat{\xi}^{\hat{g}}(\hat{g}) = \Phi(\hat{g})$ . Then, using the bounds on  $h'_{ij}(g_{ij})$  we see that

$$\hat{w}_{ij}^{\hat{g}} \in [(1/k)w_{ij}, kw_{ij}],$$

$$\Phi(g) \in [(1/k)\hat{\xi}^{\hat{g}}(g), k\hat{\xi}^{\hat{g}}(g)]$$

and

$$\Phi(g)/\Phi(g^*) \geq (1/k^2)\hat{\xi}^{\hat{g}}(g)/\hat{\xi}^{\hat{g}}(g^*).$$

This will allow us to use results from the linear case to analyze some static properties of our algorithm.

## 9. ACCURACY

In the linear case, most spectral methods consider the accuracy in solution [3]. Thus we will say that a solution is  $\epsilon$ -accurate if

$$\frac{\|x - x^*\|_{L^{q^*}}}{\|x^*\|_{L^{q^*}}} \leq \epsilon,$$

where  $L^{q^*}$  is the linearization of  $L$  at  $q^*$ , as used in the previous section. By combining our bounds on the linearization of the energy function (\*\*\*) with Lemma 6.2 in [3] for the linearization we can easily prove the following:

LEMMA 5. *If*

$$\Phi(g)/\Phi(g^*) \leq 1 + \epsilon^2/(\tau(T)k^4)$$

*then*

$$\frac{\|x - x^*\|_{L^{q^*}}}{\|x^*\|_{L^{q^*}}} \leq \epsilon$$

*where  $x$  is defined by its tree values on  $T$ .*

PROOF. It is easy to see from Lemma 6.2 in [3] that if  $\Phi(g)/\Phi(g^*) \leq (1 + \epsilon^2/k^4)$  then  $\|x - x^*\|_{\hat{L}^g} \leq \epsilon/(k^2\tau(T))$ . To complete the proof we need the following lemma relating  $\|v\|_{\hat{L}^g}$  and  $\|v\|_{\hat{L}^{g^*}}$ .

LEMMA 6.  $\|v\|_{\hat{L}^g} \leq k^2\|v\|_{\hat{L}^{g^*}}$ .

PROOF. This follows from the definition:

$$\|v\|_{\hat{L}(g)} = \sum_{(i,j) \in E} \hat{w}^g(v_i - v_j)^2$$

and noting that  $\hat{w}_{ij}^g \in [(1/k)w_{ij}, kw_{ij}]$ .  $\square$

## 10. INITIALIZATION

To initiate our algorithm we start with a b-feasible flow  $g^0$  with support only on the spanning tree. This can be computed recursively with  $O(n)$  operations by noting that for any leaf node  $i$  of the tree the value of  $g_{ij}$  for  $j = N(i)$  is given by  $w_{ij}h_{ij}(g_{ij}) = b_i$ .

However, it will be useful to note that this can be computed by first computing the solution for the linearization of  $L$  denoted  $\bar{g}^0$ , where we set all  $h_{ij}(\cdot)$  to be identity functions, using the above recursive algorithm. Then convert  $\bar{g}^0$  to  $g^0$  using the relationship  $g_{ij}^0 = h_{ij}^{-1}(\bar{g}_{ij}^0)$ .

The following bound on the initial energy follows directly from Lemma 6.1 in [3].

LEMMA 7.  $\Phi(g^0) \leq k^2\Phi(g^*)st(T)$

PROOF. Lemma 6.1 in [3] shows that  $\xi(\bar{g}^0) \leq k^2\xi(\bar{g}^*)st(T)$ . The result follows from the fact that

$$\Phi(g) \in [(1/k)\hat{\xi}^{\hat{g}}(g), k\hat{\xi}^{\hat{g}}(g)].$$

$\square$

## 11. ALGORITHM AND MAIN RESULT

We now present the algorithm which is a slight modification of that in Kelner et. al [3].

### Algorithm: Simple Nonlinear Solver

1. INPUT: Input  $G = (V, E, w)$ ,  $\epsilon \in \mathbb{R}^+$ ,  $k \geq 1$ ,  $h_{ij}(\cdot) \forall (i, j) \in E$ .
2. OUTPUT:  $g \in \mathbb{R}^E$ ,  $x \in \mathbb{R}^V$ .
3. Construct  $T$ : a low stretch spanning tree of  $G$ .
4. Construct  $g^0$ : a b-feasible solution with support only on the spanning tree.
5. Set probabilities  $p_{ij} = w_{ij}/(W_{C_{ij}}\tau(T))$  for all  $(i, j) \in E \setminus T$ .
6. Set  $S = \lceil 2k^2\tau(T)\log(k^6st(T)\tau(T)/\epsilon^2) \rceil$ .
7. For  $\ell = 1$  to  $S$ :
  - (a) Pick a random  $(i, j) \in E \setminus T$  with probability  $p_{ij}$ .
  - (b) Apply a cycle update to  $g^{\ell-1}$  using tree cycle  $C_{ij}$  to get  $g^\ell$ . Compute this to relative accuracy  $1/(2k^2)$  using binary search.
8. RETURN  $g$  and its tree induced  $x$ .

In order to show the correctness of this algorithm, we extend the analysis in [3]. Since the  $GAP(g^\ell)$  conditional on  $GAP(g^{\ell-1})$  is a random variable we focus on expected values.

LEMMA 8. *For each iteration in the loop of “Algorithm: Simple Nonlinear Solver”*

$$\frac{E[GAP(g^\ell)] - E[GAP(g^{\ell+1})|GAP(g^\ell)]}{E[GAP(g^\ell)]} \geq \frac{1}{2k^2\tau(T)}.$$

PROOF. First note that the change in GAP is the same as the change in  $\Phi(g)$  which is greater than  $W_{C_{ij}}G_{C_{ij}}^2/(2k)$  for an update of  $C_{ij}$ , thus the expected decrease is

$$E[GAP(g^\ell)] - E[GAP(g^{\ell+1})|GAP(g^\ell)] \geq \sum_{(i,j) \in E \setminus T} p_{ij}W_{C_{ij}}G_{C_{ij}}^2/(2k).$$

Substituting in  $p_{ij} = w_{ij}/(\tau(T)W_{C_{ij}})$  yields

$$E[GAP(g^\ell)] - E[GAP(g^{\ell+1})|GAP(g^\ell)] \geq \sum_{(i,j) \in E \setminus T} (w_{ij}/\tau(T))G_{C_{ij}}^2/(2k)$$

using Lemma 4 for the RHS completes the proof.  $\square$

Using this we can prove our main theorem.

THEOREM 3. *Given a nonlinear graph Laplacian  $L$ , “Algorithm: Simple Nonlinear Solver”, computes an expected  $\epsilon$ -accurate solution of  $Lx = b$  in  $\tilde{O}(mnk^2 \log(kn/\epsilon))$  time.*

PROOF. Rewriting the expression in Lemma 5 shows that in order to attain  $\epsilon$ -accuracy we need

$$\Phi(g^S) - \Phi(g^*) \leq \epsilon^2 / (\tau(T)k^4)\Phi(g^*).$$

Iterating Lemma 8, implies that

$$\Phi(g^S) - \Phi(g^*) \leq (1 - \frac{1}{2k^2\tau(T)})^S (\Phi(g^0) - \Phi(g^*)).$$

Combining this with Lemma 7 on the initial energy yields

$$\Phi(g^S) - \Phi(g^*) \leq (1 - \frac{1}{2k^2\tau(T)})^S (k^2 st(T) - 1)\Phi(g^*).$$

Plugging in  $S = \lceil 2k^2\tau(T) \log(k^6 st(T)\tau(T)/\epsilon^2) \rceil$  satisfies the requirement, proving the accuracy of the Algorithm. By Theorem 2

$$st(T) = O(m \log n \log \log n) = \tilde{O}(m)$$

the required number of iterations is given by

$$r = \tilde{O}(k^2 m \log(kn/\epsilon)).$$

Since each iteration takes  $\tilde{O}(n)$  time by Lemma 3 we get the stated result.  $\square$

Note that instead of an expected  $\epsilon$ -accurate solution, one can compute an actual  $\epsilon$ -accurate solution with high probability with the same time bound using standard probabilistic techniques.

## 12. CONCLUDING REMARKS

This paper has shown that one can extend some of the techniques from the linear theory of systems of equations arising from linear graph Laplacians to ones with nonlinear graph Laplacians. We view our analysis as suggestive that even more general extensions, perhaps even asymmetric ones related to directed graphs, may be possible, extending the range of spectral analysis. Perhaps as in Kelner et. al [3] our algorithm has a representation as an alternating projections algorithm, showing that nonlinear extensions are also possible in that setting.

There are many open directions for improving our algorithm. First, note that we have not optimized the  $k$  dependence. Next, note that one could use the ideas of tree scaling and a randomized stopping time as in [3] to reduce the running time. In addition, one might be able to use algorithms for the linear problem to speed up the nonlinear algorithm, either by providing a warm start or by speeding up convergence near optimality when the linear problem might provide a good approximation for the nonlinear one.

While this paper raises many open problems relating to nonlinear spectral theory, the main problem with the current algorithm is the dependence on  $n$  due to the complexity of computing a nonlinear cycle update which unlike the linear version does not have a simple analytic solution which can be computed in logarithmic time using a special data structure for keeping track of the flows on the tree. Nonetheless, it might be possible to reduce this complexity either through a specially designed data structure for the spanning tree, or perhaps through the use of a spanning tree with small diameter. Also, the direct application of our algorithm to nonlinear graph problems might be fruitful.

## 13. ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grants No. 1216073 and 1161813.

## 14. REFERENCES

- [1] I. Abraham and O. Neiman. Using petal-decompositions to build a low stretch spanning tree. In *Proceedings of the forty-fourth annual ACM symposium on Theory of computing*, pages 395–406. ACM, 2012.
- [2] B. Bollobás. *Modern graph theory*, volume 184. Springer Science & Business Media, 1998.
- [3] J. A. Kelner, L. Orecchia, A. Sidford, and Z. A. Zhu. A simple, combinatorial algorithm for solving sdd systems in nearly-linear time. In *Proceedings of the forty-fifth annual ACM symposium on Theory of computing*, pages 911–920. ACM, 2013.
- [4] I. Koutis, G. L. Miller, and R. Peng. A fast solver for a class of linear systems. *Communications of the ACM*, 55(10):99–107, 2012.
- [5] J. H. Reif. Efficient approximate solution of sparse linear systems. *Computers & Mathematics with Applications*, 36(9):37–58, 1998.
- [6] D. A. Spielman. Algorithms, graph theory, and linear equations in laplacian matrices. In *Proceedings of the International Congress of Mathematicians*, volume 4, pages 2698–2722, 2010.
- [7] D. A. Spielman and N. Srivastava. Graph sparsification by effective resistances. *SIAM Journal on Computing*, 40(6):1913–1926, 2011.
- [8] D. A. Spielman and S.-H. Teng. Nearly-linear time algorithms for graph partitioning, graph sparsification, and solving linear systems. In *Proceedings of the thirty-sixth annual ACM symposium on Theory of computing*, pages 81–90. ACM, 2004.
- [9] D. A. Spielman and J. Woo. A note on preconditioning by low-stretch spanning trees. *arXiv preprint arXiv:0903.2816*, 2009.
- [10] P. M. Vaidya. Solving linear equations with symmetric diagonally dominant matrices by constructing good preconditioners. A talk based on this manuscript, 1991.
- [11] N. K. Vishnoi. Laplacian solvers and their algorithmic applications. *Theoretical Computer Science*, 8(1-2):1–141, 2012.